1.8 Vector-Valued Functions

Now that we are familiar with vectors and their operations, we can begin discussing functions whose values are vectors.

Definition 1.10. A **vector-valued function of a real variable** is a rule that associates a vector $\mathbf{f}(t)$ with a real number t, where t is in some subset D of $\mathbb{R}^1$ (called the **domain** of $\mathbf{f}$). We write $\mathbf{f} : D \to \mathbb{R}^3$ to denote that $\mathbf{f}$ is a mapping of D into $\mathbb{R}^3$.

For example, $\mathbf{f}(t) = t\mathbf{i} + t^2\mathbf{j} + t^3\mathbf{k}$ is a vector-valued function in $\mathbb{R}^3$, defined for all real numbers t. We would write $\mathbf{f} : \mathbb{R} \to \mathbb{R}^3$. At $t = 1$ the value of the function is the vector $\mathbf{i} + \mathbf{j} + \mathbf{k}$, which in Cartesian coordinates has the terminal point $(1,1,1)$.

A vector-valued function of a real variable can be written in component form as

$$\mathbf{f}(t) = f_1(t)\mathbf{i} + f_2(t)\mathbf{j} + f_3(t)\mathbf{k}$$

or in the form

$$\mathbf{f}(t) = (f_1(t), f_2(t), f_3(t))$$

for some real-valued functions $f_1(t)$, $f_2(t)$, $f_3(t)$, called the *component functions* of $\mathbf{f}$. The first form is often used when emphasizing that $\mathbf{f}(t)$ is a vector, and the second form is useful when considering just the terminal points of the vectors. By identifying vectors with their terminal points, a curve in space can be written as a vector-valued function.

Example 1.35. Define $\mathbf{f} : \mathbb{R} \to \mathbb{R}^3$ by $\mathbf{f}(t) = (\cos t, \sin t, t)$.
This is the equation of a *helix* (see Figure 1.8.1). As the value of t increases, the terminal points of $\mathbf{f}(t)$ trace out a curve spiraling upward. For each t, the x- and y-coordinates of $\mathbf{f}(t)$ are $x = \cos t$ and $y = \sin t$, so

$$x^2 + y^2 = \cos^2 t + \sin^2 t = 1.$$

Thus, the curve lies on the surface of the right circular cylinder $x^2 + y^2 = 1$.

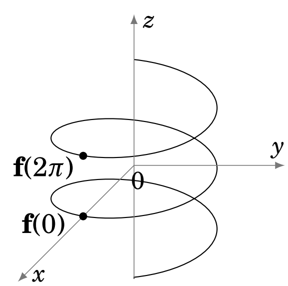


Figure 1.8.1

It may help to think of vector-valued functions of a real variable in $\mathbb{R}^3$ as a generalization of the parametric functions in $\mathbb{R}^2$ which you learned about in single-variable calculus. Much of the theory of real-valued functions of a single real variable can be applied to vector-valued functions of a real variable. Since each of the three component functions are real-valued, it will sometimes be the case that results from single-variable calculus can simply be applied to each of the component functions to yield a similar result for the vector-valued function. However, there are times when such generalizations do not hold (see Exercise 13). The concept of a limit, though, can be extended naturally to vector-valued functions, as in the following definition.

Definition 1.11. Let $\mathbf{f}(t)$ be a vector-valued function, let a be a real number and let $\mathbf{c}$ be a vector. Then we say that the **limit** of $\mathbf{f}(t)$ as t approaches a equals $\mathbf{c}$, written as $\lim_{t\to a}\mathbf{f}(t) = \mathbf{c}$, if $\lim_{t\to a} \|\mathbf{f}(t) - \mathbf{c}\| = 0$. If $\mathbf{f}(t) = (f_1(t), f_2(t), f_3(t))$, then

$$\lim_{t\to a}\mathbf{f}(t) = \left(\lim_{t\to a} f_1(t), \lim_{t\to a} f_2(t), \lim_{t\to a} f_3(t)\right)$$

provided that all three limits on the right side exist.

The above definition shows that continuity and the derivative of vector-valued functions can also be defined in terms of its component functions.

Definition 1.12. Let $\mathbf{f}(t) = (f_1(t), f_2(t), f_3(t))$ be a vector-valued function, and let a be a real number in its domain. Then $\mathbf{f}(t)$ is **continuous** at a if $\lim_{t\to a}\mathbf{f}(t) = \mathbf{f}(a)$. Equivalently, $\mathbf{f}(t)$ is continuous at a if and only if $f_1(t)$, $f_2(t)$, and $f_3(t)$ are continuous at a.
The **derivative** of $\mathbf{f}(t)$ at a, denoted by $\mathbf{f}'(a)$ or $\frac{d\mathbf{f}}{dt}(a)$, is the limit

$$\mathbf{f}'(a) = \lim_{h\to 0} \frac{\mathbf{f}(a+h) - \mathbf{f}(a)}{h}$$

if that limit exists. Equivalently, $\mathbf{f}'(a) = (f_1{}'(a), f_2{}'(a), f_3{}'(a))$, if the component derivatives exist. We say that $\mathbf{f}(t)$ is **differentiable** at a if $\mathbf{f}'(a)$ exists.

Recall that the derivative of a real-valued function of a single variable is a real number, representing the slope of the tangent line to the graph of the function at a point. Similarly, the derivative of a vector-valued function is a **tangent vector** to the curve in space which the function represents, and it lies on the *tangent line* to the curve (see Figure 1.8.2).

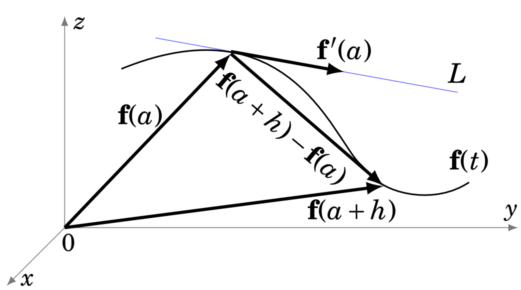


Figure 1.8.2 Tangent vector $\mathbf{f}'(a)$ and tangent line $L = \mathbf{f}(a) + s\mathbf{f}'(a)$

Example 1.36. Let $\mathbf{f}(t) = (\cos t, \sin t, t)$. Then $\mathbf{f}'(t) = (-\sin t, \cos t, 1)$ for all t. The tangent line L to the curve at $\mathbf{f}(2\pi) = (1, 0, 2\pi)$ is $L = \mathbf{f}(2\pi) + s\,\mathbf{f}'(2\pi) = (1, 0, 2\pi) + s(0, 1, 1)$, or in parametric form: $x = 1$, $y = s$, $z = 2\pi + s$ for $-\infty < s < \infty$.

A **scalar function** is a real-valued function. Note that if $u(t)$ is a scalar function and $\mathbf{f}(t)$ is a vector-valued function, then their product, defined by $(u\,\mathbf{f})(t) = u(t)\mathbf{f}(t)$ for all t, is a vector-valued function (since the product of a scalar with a vector is a vector).

The basic properties of derivatives of vector-valued functions are summarized in the following theorem.

Theorem 1.20. Let $\mathbf{f}(t)$ and $\mathbf{g}(t)$ be differentiable vector-valued functions, let $u(t)$ be a differentiable scalar function, let k be a scalar, and let $\mathbf{c}$ be a constant vector. Then

(a) $\dfrac{d}{dt}(\mathbf{c}) = \mathbf{0}$

(b) $\dfrac{d}{dt}(k\mathbf{f}) = k\dfrac{d\mathbf{f}}{dt}$

(c) $\dfrac{d}{dt}(\mathbf{f}+\mathbf{g}) = \dfrac{d\mathbf{f}}{dt} + \dfrac{d\mathbf{g}}{dt}$

(d) $\dfrac{d}{dt}(\mathbf{f}-\mathbf{g}) = \dfrac{d\mathbf{f}}{dt} - \dfrac{d\mathbf{g}}{dt}$

(e) $\dfrac{d}{dt}(u\,\mathbf{f}) = \dfrac{du}{dt}\mathbf{f} + u\dfrac{d\mathbf{f}}{dt}$

(f) $\dfrac{d}{dt}(\mathbf{f}\cdot\mathbf{g}) = \dfrac{d\mathbf{f}}{dt}\cdot\mathbf{g} + \mathbf{f}\cdot\dfrac{d\mathbf{g}}{dt}$

(g) $\dfrac{d}{dt}(\mathbf{f}\times\mathbf{g}) = \dfrac{d\mathbf{f}}{dt}\times\mathbf{g} + \mathbf{f}\times\dfrac{d\mathbf{g}}{dt}$

Proof: The proofs of parts (a)-(e) follow easily by differentiating the component functions and using the rules for derivatives from single-variable calculus. We will prove part (f), and leave the proof of part (g) as an exercise for the reader.

(f) Write $\mathbf{f}(t) = (f_1(t), f_2(t), f_3(t))$ and $\mathbf{g}(t) = (g_1(t), g_2(t), g_3(t))$, where the component functions $f_1(t)$, $f_2(t)$, $f_3(t)$, $g_1(t)$, $g_2(t)$, $g_3(t)$ are all differentiable real-valued functions. Then

$$\begin{aligned}
\frac{d}{dt}(\mathbf{f}(t)\cdot\mathbf{g}(t)) &= \frac{d}{dt}(f_1(t)g_1(t) + f_2(t)g_2(t) + f_3(t)g_3(t)) \\
&= \frac{d}{dt}(f_1(t)g_1(t)) + \frac{d}{dt}(f_2(t)g_2(t)) + \frac{d}{dt}(f_3(t)g_3(t)) \\
&= \frac{df_1}{dt}(t)g_1(t) + f_1(t)\frac{dg_1}{dt}(t) + \frac{df_2}{dt}(t)g_2(t) + f_2(t)\frac{dg_2}{dt}(t) + \frac{df_3}{dt}(t)g_3(t) + f_3(t)\frac{dg_3}{dt}(t) \\
&= \left(\frac{df_1}{dt}(t), \frac{df_2}{dt}(t), \frac{df_3}{dt}(t)\right)\cdot(g_1(t), g_2(t), g_3(t)) \\
&\quad + (f_1(t), f_2(t), f_3(t))\cdot\left(\frac{dg_1}{dt}(t), \frac{dg_2}{dt}(t), \frac{dg_3}{dt}(t)\right) \\
&= \frac{d\mathbf{f}}{dt}(t)\cdot\mathbf{g}(t) + \mathbf{f}(t)\cdot\frac{d\mathbf{g}}{dt}(t) \ \text{ for all } t.
\end{aligned}$$

QED

Example 1.37. Suppose $\mathbf{f}(t)$ is differentiable. Find the derivative of $\|\mathbf{f}(t)\|$.

Solution: Since $\|\mathbf{f}(t)\|$ is a real-valued function of t, then by the Chain Rule for real-valued functions, we know that $\frac{d}{dt}\|\mathbf{f}(t)\|^2 = 2\|\mathbf{f}(t)\|\frac{d}{dt}\|\mathbf{f}(t)\|$.

But $\|\mathbf{f}(t)\|^2 = \mathbf{f}(t)\cdot\mathbf{f}(t)$, so $\frac{d}{dt}\|\mathbf{f}(t)\|^2 = \frac{d}{dt}(\mathbf{f}(t)\cdot\mathbf{f}(t))$. Hence, we have

$$\begin{aligned} 2\|\mathbf{f}(t)\|\frac{d}{dt}\|\mathbf{f}(t)\| &= \frac{d}{dt}(\mathbf{f}(t)\cdot\mathbf{f}(t)) = \mathbf{f}'(t)\cdot\mathbf{f}(t) + \mathbf{f}(t)\cdot\mathbf{f}'(t) \text{ by Theorem 1.20(f), so} \\ &= 2\mathbf{f}'(t)\cdot\mathbf{f}(t)\text{ , so if } \|\mathbf{f}(t)\| \neq 0 \text{ then} \\ \frac{d}{dt}\|\mathbf{f}(t)\| &= \frac{\mathbf{f}'(t)\cdot\mathbf{f}(t)}{\|\mathbf{f}(t)\|}. \end{aligned}$$

We know that $\|\mathbf{f}(t)\|$ is constant if and only if $\frac{d}{dt}\|\mathbf{f}(t)\| = 0$ for all t. Also, $\mathbf{f}(t)\perp\mathbf{f}'(t)$ if and only if $\mathbf{f}'(t)\cdot\mathbf{f}(t) = 0$. Thus, the above example shows this important fact:

If $\|\mathbf{f}(t)\| \neq 0$, then $\|\mathbf{f}(t)\|$ is constant if and only if $\mathbf{f}(t)\perp\mathbf{f}'(t)$ for all t.

This means that if a curve lies completely on a sphere (or circle) centered at the origin, then the tangent vector $\mathbf{f}'(t)$ is always perpendicular to the *position vector* $\mathbf{f}(t)$.

Example 1.38. The *spherical spiral* $\mathbf{f}(t) = \left(\frac{\cos t}{\sqrt{1+a^2t^2}}, \frac{\sin t}{\sqrt{1+a^2t^2}}, \frac{-at}{\sqrt{1+a^2t^2}}\right)$, for $a \neq 0$.

Figure 1.8.3 shows the graph of the curve when $a = 0.2$. In the exercises, the reader will be asked to show that this curve lies on the sphere $x^2 + y^2 + z^2 = 1$ and to verify directly that $\mathbf{f}'(t)\cdot\mathbf{f}(t) = 0$ for all t.

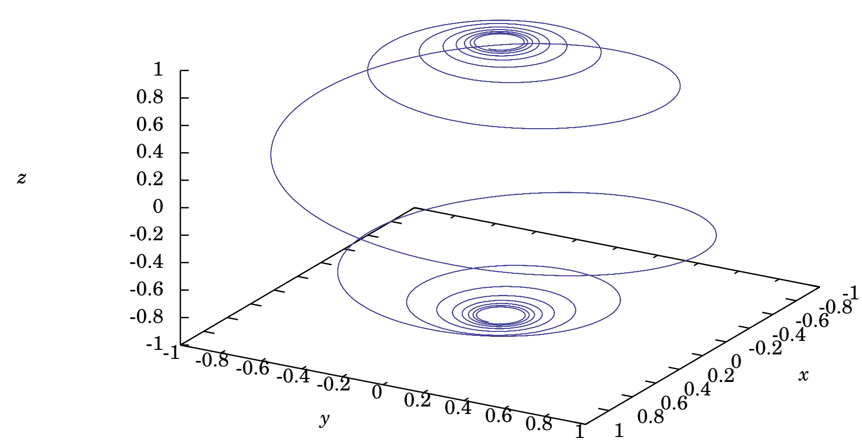


Figure 1.8.3 Spherical spiral with $a = 0.2$

Just as in single-variable calculus, higher-order derivatives of vector-valued functions are obtained by repeatedly differentiating the (first) derivative of the function:

$$\mathbf{f}''(t) = \frac{d}{dt}\mathbf{f}'(t)\,, \qquad \mathbf{f}'''(t) = \frac{d}{dt}\mathbf{f}''(t)\,, \quad \ldots \quad , \quad \frac{d^n\mathbf{f}}{dt^n} = \frac{d}{dt}\left(\frac{d^{n-1}\mathbf{f}}{dt^{n-1}}\right) \quad (\text{for } n = 2,3,4,\ldots)$$

We can use vector-valued functions to represent physical quantities, such as velocity, acceleration, force, momentum, etc. For example, let the real variable t represent time elapsed from some initial time ($t = 0$), and suppose that an object of constant mass m is subjected to some force so that it moves in space, with its position (x, y, z) at time t a function of t. That is, $x = x(t)$, $y = y(t)$, $z = z(t)$ for some real-valued functions $x(t)$, $y(t)$, $z(t)$. Call $\mathbf{r}(t) = (x(t), y(t), z(t))$ the **position vector** of the object. We can define various physical quantities associated with the object as follows:[14]

$$\begin{aligned}
\textit{position}\colon\ \mathbf{r}(t) &= (x(t), y(t), z(t)) \\
\textit{velocity}\colon\ \mathbf{v}(t) &= \dot{\mathbf{r}}(t) = \mathbf{r}'(t) = \frac{d\mathbf{r}}{dt} \\
&= (x'(t), y'(t), z'(t)) \\
\textit{acceleration}\colon\ \mathbf{a}(t) &= \dot{\mathbf{v}}(t) = \mathbf{v}'(t) = \frac{d\mathbf{v}}{dt} \\
&= \ddot{\mathbf{r}}(t) = \mathbf{r}''(t) = \frac{d^2\mathbf{r}}{dt^2} \\
&= (x''(t), y''(t), z''(t)) \\
\textit{momentum}\colon\ \mathbf{p}(t) &= m\mathbf{v}(t) \\
\textit{force}\colon\ \mathbf{F}(t) &= \dot{\mathbf{p}}(t) = \mathbf{p}'(t) = \frac{d\mathbf{p}}{dt} \quad \text{(Newton's Second Law of Motion)}
\end{aligned}$$

The magnitude $\|\mathbf{v}(t)\|$ of the velocity vector is called the *speed* of the object. Note that since the mass m is a constant, the force equation becomes the familiar $\mathbf{F}(t) = m\mathbf{a}(t)$.

Example 1.39. Let $\mathbf{r}(t) = (5\cos t, 3\sin t, 4\sin t)$ be the position vector of an object at time $t \ge 0$. Find its (a) velocity and (b) acceleration vectors.

Solution: (a) $\mathbf{v}(t) = \dot{\mathbf{r}}(t) = (-5\sin t, 3\cos t, 4\cos t)$

(b) $\mathbf{a}(t) = \dot{\mathbf{v}}(t) = (-5\cos t, -3\sin t, -4\sin t)$

Note that $\|\mathbf{r}(t)\| = \sqrt{25\cos^2 t + 25\sin^2 t} = 5$ for all t, so by Example 1.37 we know that $\mathbf{r}(t) \cdot \dot{\mathbf{r}}(t) = 0$ for all t (which we can verify from part (a)). In fact, $\|\mathbf{v}(t)\| = 5$ for all t also. And not only does $\mathbf{r}(t)$ lie on the sphere of radius 5 centered at the origin, but perhaps not so obvious is that it lies completely within a *circle* of radius 5 centered at the origin. Also, note that $\mathbf{a}(t) = -\mathbf{r}(t)$. It turns out (see Exercise 16) that whenever an object moves in a circle with constant speed, the acceleration vector will point in the opposite direction of the position vector (i.e. towards the center of the circle).

[14] We will often use the older dot notation for derivatives when physics is involved.

Recall from Section 1.5 that if $\mathbf{r}_1$, $\mathbf{r}_2$ are position vectors to distinct points then $\mathbf{r}_1 + t(\mathbf{r}_2 - \mathbf{r}_1)$ represents a line through those two points as t varies over all real numbers. That vector sum can be written as $(1-t)\mathbf{r}_1 + t\mathbf{r}_2$. So the function $\mathbf{l}(t) = (1-t)\mathbf{r}_1 + t\mathbf{r}_2$ is a line through the terminal points of $\mathbf{r}_1$ and $\mathbf{r}_2$, and when t is restricted to the interval $[0,1]$ it is the line segment between the points, with $\mathbf{l}(0) = \mathbf{r}_1$ and $\mathbf{l}(1) = \mathbf{r}_2$.

In general, a function of the form $\mathbf{f}(t) = (a_1 t + b_1, a_2 t + b_2, a_3 t + b_3)$ represents a line in $\mathbb{R}^3$. A function of the form $\mathbf{f}(t) = (a_1 t^2 + b_1 t + c_1, a_2 t^2 + b_2 t + c_2, a_3 t^2 + b_3 t + c_3)$ represents a (possibly degenerate) parabola in $\mathbb{R}^3$.

Example 1.40. *Bézier curves* are used in Computer Aided Design (CAD) to approximate the shape of a polygonal path in space (called the *Bézier polygon* or *control polygon*). For instance, given three points (or position vectors) $\mathbf{b}_0$, $\mathbf{b}_1$, $\mathbf{b}_2$ in $\mathbb{R}^3$, define

$$\begin{aligned}
\mathbf{b}_0^1(t) &= (1-t)\mathbf{b}_0 + t\mathbf{b}_1 \\
\mathbf{b}_1^1(t) &= (1-t)\mathbf{b}_1 + t\mathbf{b}_2 \\
\mathbf{b}_0^2(t) &= (1-t)\mathbf{b}_0^1(t) + t\mathbf{b}_1^1(t) \\
&= (1-t)^2\mathbf{b}_0 + 2t(1-t)\mathbf{b}_1 + t^2\mathbf{b}_2
\end{aligned}$$

for all real t. For t in the interval $[0,1]$, we see that $\mathbf{b}_0^1(t)$ is the line segment between $\mathbf{b}_0$ and $\mathbf{b}_1$, and $\mathbf{b}_1^1(t)$ is the line segment between $\mathbf{b}_1$ and $\mathbf{b}_2$. The function $\mathbf{b}_0^2(t)$ is the Bézier curve for the points $\mathbf{b}_0$, $\mathbf{b}_1$, $\mathbf{b}_2$. Note from the last formula that the curve is a parabola that goes through $\mathbf{b}_0$ (when $t=0$) and $\mathbf{b}_2$ (when $t=1$).

As an example, let $\mathbf{b}_0 = (0,0,0)$, $\mathbf{b}_1 = (1,2,3)$, and $\mathbf{b}_2 = (4,5,2)$. Then the explicit formula for the Bézier curve is $\mathbf{b}_0^2(t) = (2t + 2t^2, 4t + t^2, 6t - 4t^2)$, as shown in Figure 1.8.4, where the line segments are $\mathbf{b}_0^1(t)$ and $\mathbf{b}_1^1(t)$, and the curve is $\mathbf{b}_0^2(t)$.

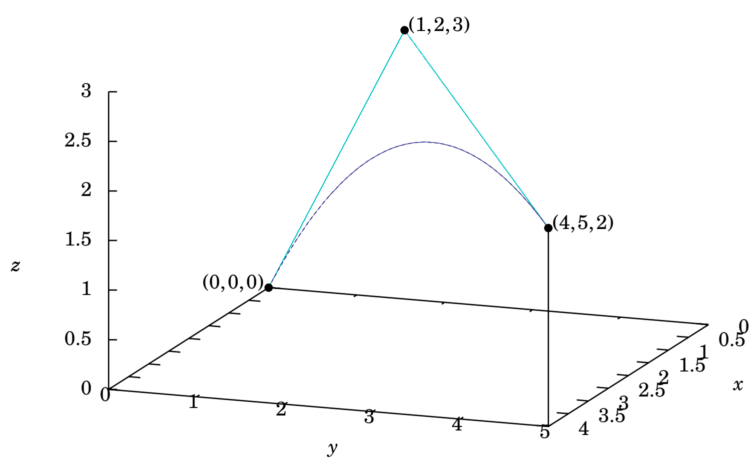


Figure 1.8.4 Bézier curve approximation for three points

In general, the polygonal path determined by $n \ge 3$ noncollinear points in $\mathbb{R}^3$ can be used to define the Bézier curve recursively by a process called *repeated linear interpolation*. This curve will be a vector-valued function whose components are polynomials of degree $n-1$, and its formula is given by *de Casteljau's algorithm*.[15] In the exercises, the reader will be given the algorithm for the case of $n = 4$ points and asked to write the explicit formula for the Bézier curve for the four points shown in Figure 1.8.5.

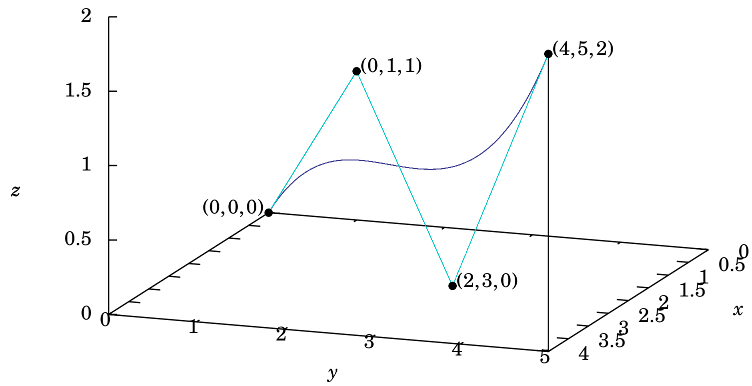


Figure 1.8.5 Bézier curve approximation for four points

Exercises

A

For Exercises 1-4, calculate $\mathbf{f}'(t)$ and find the tangent line at $\mathbf{f}(0)$.

1. $\mathbf{f}(t) = (t+1, t^2+1, t^3+1)$

2. $\mathbf{f}(t) = (e^t+1, e^{2t}+1, e^{t^2}+1)$

3. $\mathbf{f}(t) = (\cos 2t, \sin 2t, t)$

4. $\mathbf{f}(t) = (\sin 2t, 2\sin^2 t, 2\cos t)$

For Exercises 5-6, find the velocity $\mathbf{v}(t)$ and acceleration $\mathbf{a}(t)$ of an object with the given position vector $\mathbf{r}(t)$.

5. $\mathbf{r}(t) = (t, t - \sin t, 1 - \cos t)$

6. $\mathbf{r}(t) = (3\cos t, 2\sin t, 1)$

B

7. Let $\mathbf{f}(t) = \left(\dfrac{\cos t}{\sqrt{1+a^2t^2}}, \dfrac{\sin t}{\sqrt{1+a^2t^2}}, \dfrac{-at}{\sqrt{1+a^2t^2}}\right)$, with $a \ne 0$.

(a) Show that $\|\mathbf{f}(t)\| = 1$ for all t.

(b) Show directly that $\mathbf{f}'(t) \cdot \mathbf{f}(t) = 0$ for all t.

8. If $\mathbf{f}'(t) = \mathbf{0}$ for all t in some interval (a, b), show that $\mathbf{f}(t)$ is a constant vector in (a, b).

[15] See pp. 27-30 in FARIN.

9. For a constant vector $\mathbf{c} \neq \mathbf{0}$, the function $\mathbf{f}(t) = t\mathbf{c}$ represents a line parallel to $\mathbf{c}$.

(a) What kind of curve does $\mathbf{g}(t) = t^3\mathbf{c}$ represent? Explain.

(b) What kind of curve does $\mathbf{h}(t) = e^t\mathbf{c}$ represent? Explain.

(c) Compare $\mathbf{f}'(0)$ and $\mathbf{g}'(0)$. Given your answer to part (a), how do you explain the difference in the two derivatives?

10. Show that $\dfrac{d}{dt}\left(\mathbf{f} \times \dfrac{d\mathbf{f}}{dt}\right) = \mathbf{f} \times \dfrac{d^2\mathbf{f}}{dt^2}$.

11. Let a particle of (constant) mass m have position vector $\mathbf{r}(t)$, velocity $\mathbf{v}(t)$, acceleration $\mathbf{a}(t)$ and momentum $\mathbf{p}(t)$ at time t. The *angular momentum* $\mathbf{L}(t)$ of the particle with respect to the origin at time t is defined as $\mathbf{L}(t) = \mathbf{r}(t) \times \mathbf{p}(t)$. If $\mathbf{F}(t)$ is the force acting on the particle at time t, then define the *torque* $\mathbf{N}(t)$ acting on the particle with respect to the origin as $\mathbf{N}(t) = \mathbf{r}(t) \times \mathbf{F}(t)$. Show that $\mathbf{L}'(t) = \mathbf{N}(t)$.

12. Show that $\dfrac{d}{dt}(\mathbf{f} \cdot (\mathbf{g} \times \mathbf{h})) = \dfrac{d\mathbf{f}}{dt} \cdot (\mathbf{g} \times \mathbf{h}) + \mathbf{f} \cdot \left(\dfrac{d\mathbf{g}}{dt} \times \mathbf{h}\right) + \mathbf{f} \cdot \left(\mathbf{g} \times \dfrac{d\mathbf{h}}{dt}\right)$.

13. The Mean Value Theorem does not hold for vector-valued functions: Show that for $\mathbf{f}(t) = (\cos t, \sin t, t)$, there is no t in the interval $(0, 2\pi)$ such that

$$\mathbf{f}'(t) = \frac{\mathbf{f}(2\pi) - \mathbf{f}(0)}{2\pi - 0}.$$

C

14. The Bézier curve $\mathbf{b}_0^3(t)$ for four noncollinear points $\mathbf{b}_0$, $\mathbf{b}_1$, $\mathbf{b}_2$, $\mathbf{b}_3$ in $\mathbb{R}^3$ is defined by the following algorithm (going from the left column to the right):

$$\begin{array}{lll}
\mathbf{b}_0^1(t) = (1-t)\mathbf{b}_0 + t\mathbf{b}_1 & \mathbf{b}_0^2(t) = (1-t)\mathbf{b}_0^1(t) + t\mathbf{b}_1^1(t) & \mathbf{b}_0^3(t) = (1-t)\mathbf{b}_0^2(t) + t\mathbf{b}_1^2(t) \\
\mathbf{b}_1^1(t) = (1-t)\mathbf{b}_1 + t\mathbf{b}_2 & \mathbf{b}_1^2(t) = (1-t)\mathbf{b}_1^1(t) + t\mathbf{b}_2^1(t) & \\
\mathbf{b}_2^1(t) = (1-t)\mathbf{b}_2 + t\mathbf{b}_3 & &
\end{array}$$

(a) Show that $\mathbf{b}_0^3(t) = (1-t)^3\mathbf{b}_0 + 3t(1-t)^2\mathbf{b}_1 + 3t^2(1-t)\mathbf{b}_2 + t^3\mathbf{b}_3$.

(b) Write the explicit formula (as in Example 1.40) for the Bézier curve for the points $\mathbf{b}_0 = (0,0,0)$, $\mathbf{b}_1 = (0,1,1)$, $\mathbf{b}_2 = (2,3,0)$, $\mathbf{b}_3 = (4,5,2)$.

15. Let $\mathbf{r}(t)$ be the position vector for a particle moving in $\mathbb{R}^3$. Show that

$$\frac{d}{dt}(\mathbf{r} \times (\mathbf{v} \times \mathbf{r})) = \|\mathbf{r}\|^2\mathbf{a} + (\mathbf{r} \cdot \mathbf{v})\mathbf{v} - (\|\mathbf{v}\|^2 + \mathbf{r} \cdot \mathbf{a})\mathbf{r}.$$

16. Let $\mathbf{r}(t)$ be the position vector in $\mathbb{R}^3$ for a particle that moves with constant speed $c > 0$ in a circle of radius $a > 0$ in the xy-plane. Show that $\mathbf{a}(t)$ points in the opposite direction as $\mathbf{r}(t)$ for all t. (*Hint: Use Example 1.37 to show that* $\mathbf{r}(t) \perp \mathbf{v}(t)$ *and* $\mathbf{a}(t) \perp \mathbf{v}(t)$, *and hence* $\mathbf{a}(t) \parallel \mathbf{r}(t)$.)

17. Prove Theorem 1.20(g).

1.9 Arc Length

Let $\mathbf{r}(t) = (x(t), y(t), z(t))$ be the position vector of an object moving in $\mathbb{R}^3$. Since $\|\mathbf{v}(t)\|$ is the speed of the object at time t, it seems natural to define the distance s traveled by the object from time $t = a$ to $t = b$ as the definite integral

$$s = \int_a^b \|\mathbf{v}(t)\|\,dt = \int_a^b \sqrt{x'(t)^2 + y'(t)^2 + z'(t)^2}\,dt\,, \tag{1.40}$$

which is analogous to the case from single-variable calculus for parametric functions in $\mathbb{R}^2$. This is indeed how we will define the distance traveled and, in general, the arc length of a curve in $\mathbb{R}^3$.

Definition 1.13. Let $\mathbf{f}(t) = (x(t), y(t), z(t))$ be a curve in $\mathbb{R}^3$ whose domain includes the interval $[a,b]$. Suppose that in the interval (a,b) the first derivative of each component function $x(t)$, $y(t)$ and $z(t)$ exists and is continuous, and that no section of the curve is repeated. Then the **arc length** L of the curve from $t = a$ to $t = b$ is

$$L = \int_a^b \|\mathbf{f}'(t)\|\,dt = \int_a^b \sqrt{x'(t)^2 + y'(t)^2 + z'(t)^2}\,dt \tag{1.41}$$

A real-valued function whose first derivative is continuous is called *continuously differentiable* (or a $\mathscr{C}^1$ function), and a function whose derivatives of all orders are continuous is called *smooth* (or a $\mathscr{C}^\infty$ function). All the functions we will consider will be smooth. A *smooth curve* $\mathbf{f}(t)$ is one whose derivative $\mathbf{f}'(t)$ is never the zero vector and whose component functions are all smooth.

Note that we did not *prove* that the formula in the above definition actually gives the length of a section of a curve. A rigorous proof requires dealing with some subtleties, normally glossed over in calculus texts, which are beyond the scope of this book.[16]

Example 1.41. Find the length L of the helix $\mathbf{f}(t) = (\cos t, \sin t, t)$ from $t = 0$ to $t = 2\pi$.

Solution: By formula (1.41), we have

$$\begin{aligned} L &= \int_0^{2\pi} \sqrt{(-\sin t)^2 + (\cos t)^2 + 1^2}\,dt = \int_0^{2\pi} \sqrt{\sin^2 t + \cos^2 t + 1}\,dt = \int_0^{2\pi} \sqrt{2}\,dt \\ &= \sqrt{2}(2\pi - 0) = 2\sqrt{2}\pi \end{aligned}$$

Similar to the case in $\mathbb{R}^2$, if there are values of t in the interval $[a,b]$ where the derivative of a component function is not continuous then it is often possible to partition $[a,b]$ into subintervals where all the component functions are continuously differentiable (except at the endpoints, which can be ignored). The sum of the arc lengths over the subintervals will be the arc length over $[a,b]$.

[16] In particular, *Duhamel's principle* is needed. See the proof in TAYLOR and MANN, § 14.2 and § 18.2.

Notice that the curve traced out by the function $\mathbf{f}(t) = (\cos t, \sin t, t)$ from Example 1.41 is also traced out by the function $\mathbf{g}(t) = (\cos 2t, \sin 2t, 2t)$. For example, over the interval $[0,\pi]$, $\mathbf{g}(t)$ traces out the same section of the curve as $\mathbf{f}(t)$ does over the interval $[0,2\pi]$. Intuitively, this says that $\mathbf{g}(t)$ traces the curve twice as fast as $\mathbf{f}(t)$. This makes sense since, viewing the functions as position vectors and their derivatives as velocity vectors, the speeds of $\mathbf{f}(t)$ and $\mathbf{g}(t)$ are $\|\mathbf{f}'(t)\| = \sqrt{2}$ and $\|\mathbf{g}'(t)\| = 2\sqrt{2}$, respectively. We say that $\mathbf{g}(t)$ and $\mathbf{f}(t)$ are different *parametrizations* of the same curve.

Definition 1.14. Let C be a smooth curve in $\mathbb{R}^3$ represented by a function $\mathbf{f}(t)$ defined on an interval $[a,b]$, and let $\alpha : [c,d] \to [a,b]$ be a smooth one-to-one mapping of an interval $[c,d]$ onto $[a,b]$. Then the function $\mathbf{g} : [c,d] \to \mathbb{R}^3$ defined by $\mathbf{g}(s) = \mathbf{f}(\alpha(s))$ is a **parametrization** of C with **parameter** s. If α is strictly increasing on $[c,d]$ then we say that $\mathbf{g}(s)$ is *equivalent* to $\mathbf{f}(t)$.

$$\begin{tikzcd} s & t & \mathbf{f}(t) \\ {[c,d]} \arrow[r, "\alpha"] \arrow[rr, bend right, "\mathbf{g}(s) = \mathbf{f}(\alpha(s)) = \mathbf{f}(t)"'] & {[a,b]} \arrow[r, "\mathbf{f}"] & \mathbb{R}^3 \end{tikzcd}$$

Note that the differentiability of $\mathbf{g}(s)$ follows from a version of the Chain Rule for vector-valued functions (the proof is left as an exercise):

Theorem 1.21. Chain Rule: If $\mathbf{f}(t)$ is a differentiable vector-valued function of t, and $t = \alpha(s)$ is a differentiable scalar function of s, then $\mathbf{f}(s) = \mathbf{f}(\alpha(s))$ is a differentiable vector-valued function of s, and

$$\frac{d\mathbf{f}}{ds} = \frac{d\mathbf{f}}{dt}\frac{dt}{ds} \tag{1.42}$$

for any s where the composite function $\mathbf{f}(\alpha(s))$ is defined.

Example 1.42. The following are all equivalent parametrizations of the same curve:

$$\begin{aligned} \mathbf{f}(t) &= (\cos t, \sin t, t) \quad \text{for } t \text{ in } [0,2\pi] \\ \mathbf{g}(s) &= (\cos 2s, \sin 2s, 2s) \quad \text{for } s \text{ in } [0,\pi] \\ \mathbf{h}(s) &= (\cos 2\pi s, \sin 2\pi s, 2\pi s) \quad \text{for } s \text{ in } [0,1] \end{aligned}$$

To see that $\mathbf{g}(s)$ is equivalent to $\mathbf{f}(t)$, define $\alpha : [0,\pi] \to [0,2\pi]$ by $\alpha(s) = 2s$. Then α is smooth, one-to-one, maps $[0,\pi]$ onto $[0,2\pi]$, and is strictly increasing (since $\alpha'(s) = 2 > 0$ for all s). Likewise, defining $\alpha : [0,1] \to [0,2\pi]$ by $\alpha(s) = 2\pi s$ shows that $\mathbf{h}(s)$ is equivalent to $\mathbf{f}(t)$.

A curve can have many parametrizations, with different speeds, so which one is the best to use? In some situations the **arc length parametrization** can be useful. The idea behind this is to replace the parameter t, for any given smooth parametrization $\mathbf{f}(t)$ defined on $[a,b]$, by the parameter s given by

$$s = s(t) = \int_a^t \|\mathbf{f}'(u)\|\,du. \tag{1.43}$$

In terms of motion along a curve, s is the distance traveled along the curve after time t has elapsed. So the new parameter will be distance instead of time. There is a natural correspondence between s and t: from a starting point on the curve, the distance traveled along the curve (in one direction) is uniquely determined by the amount of time elapsed, and vice versa.

Since s is the arc length of the curve over the interval $[a,t]$ for each t in $[a,b]$, then it is a function of t. By the Fundamental Theorem of Calculus, its derivative is

$$s'(t) = \frac{ds}{dt} = \frac{d}{dt}\int_a^t \|\mathbf{f}'(u)\|\,du = \|\mathbf{f}'(t)\| \quad \text{for all } t \text{ in } [a,b].$$

Since $\mathbf{f}(t)$ is smooth, then $\|\mathbf{f}'(t)\| > 0$ for all t in $[a,b]$. Thus $s'(t) > 0$ and hence $s(t)$ is strictly increasing on the interval $[a,b]$. Recall that this means that s is a one-to-one mapping of the interval $[a,b]$ onto the interval $[s(a), s(b)]$. But we see that

$$s(a) = \int_a^a \|\mathbf{f}'(u)\|\,du = 0 \quad \text{and} \quad s(b) = \int_a^b \|\mathbf{f}'(u)\|\,du = L = \text{arc length from } t = a \text{ to } t = b$$

So the function $s : [a,b] \to [0,L]$ is a one-to-one, differentiable mapping onto the interval $[0,L]$. From single-variable calculus, we know that this means that there exists an inverse function $\alpha : [0,L] \to [a,b]$ that is differentiable and the inverse of $s : [a,b] \to [0,L]$. That is, for each t in $[a,b]$ there is a unique s in $[0,L]$ such that $s = s(t)$ and $t = \alpha(s)$. And we know that the derivative of α is

$$\begin{tikzcd} \underset{[0,L]}{s} \arrow[rr, bend left, "\alpha(s)"] & & \underset{[a,b]}{t} \arrow[ll, bend left, "s(t)"] \end{tikzcd}$$

Figure 1.9.1 $t = \alpha(s)$

$$\alpha'(s) = \frac{1}{s'(\alpha(s))} = \frac{1}{\|\mathbf{f}'(\alpha(s))\|}$$

So define the arc length parametrization $\mathbf{f} : [0,L] \to \mathbb{R}^3$ by

$$\mathbf{f}(s) = \mathbf{f}(\alpha(s)) \quad \text{for all } s \text{ in } [0,L].$$

Then $\mathbf{f}(s)$ is smooth, by the Chain Rule. In fact, $\mathbf{f}(s)$ has *unit speed*:

$$\begin{aligned} \mathbf{f}'(s) &= \mathbf{f}'(\alpha(s))\,\alpha'(s) \quad \text{by the Chain Rule, so} \\ &= \mathbf{f}'(\alpha(s))\,\frac{1}{\|\mathbf{f}'(\alpha(s))\|}\ , \text{ so} \\ \|\mathbf{f}'(s)\| &= 1 \quad \text{for all } s \text{ in } [0,L]. \end{aligned}$$

So the arc length parametrization traverses the curve at a “normal” rate.

In practice, parametrizing a curve $\mathbf{f}(t)$ by arc length requires you to evaluate the integral $s = \int_a^t \|\mathbf{f}'(u)\|\,du$ in some closed form (as a function of t) so that you could then solve for t in terms of s. If that can be done, you would then substitute the expression for t in terms of s (which we called $\alpha(s)$) into the formula for $\mathbf{f}(t)$ to get $\mathbf{f}(s)$.

Example 1.43. Parametrize the helix $\mathbf{f}(t) = (\cos t, \sin t, t)$, for t in $[0, 2\pi]$, by arc length.

Solution: By Example 1.41 and formula (1.43), we have

$$s = \int_0^t \|\mathbf{f}'(u)\|\,du = \int_0^t \sqrt{2}\,du = \sqrt{2}\,t \quad \text{for all } t \text{ in } [0, 2\pi].$$

So we can solve for t in terms of s: $t = \alpha(s) = \dfrac{s}{\sqrt{2}}$.

$\therefore\ \ \mathbf{f}(s) = \left(\cos \dfrac{s}{\sqrt{2}}, \sin \dfrac{s}{\sqrt{2}}, \dfrac{s}{\sqrt{2}}\right)$ for all s in $[0, 2\sqrt{2}\pi]$. Note that $\|\mathbf{f}'(s)\| = 1$.

Arc length plays an important role when discussing *curvature* and *moving frame fields*, in the field of mathematics known as *differential geometry*.[17] The methods involve using an arc length parametrization, which often leads to an integral that is either difficult or impossible to evaluate in a simple closed form. The simple integral in Example 1.43 is the exception, not the norm. In general, arc length parametrizations are more useful for theoretical purposes than for practical computations.[18] Curvature and moving frame fields can be defined without using arc length, which makes their computation much easier, and these definitions can be shown to be equivalent to those using arc length. We will leave this to the exercises.

The arc length for curves given in other coordinate systems can also be calculated:

Theorem 1.22. Suppose that $r = r(t)$, $\theta = \theta(t)$ and $z = z(t)$ are the cylindrical coordinates of a curve $\mathbf{f}(t)$, for t in $[a, b]$. Then the arc length L of the curve over $[a, b]$ is

$$L = \int_a^b \sqrt{r'(t)^2 + r(t)^2\theta'(t)^2 + z'(t)^2}\,dt \tag{1.44}$$

Proof: The Cartesian coordinates $(x(t), y(t), z(t))$ of a point on the curve are given by

$$x(t) = r(t)\cos\theta(t), \quad y(t) = r(t)\sin\theta(t), \quad z(t) = z(t)$$

so differentiating the above expressions for $x(t)$ and $y(t)$ with respect to t gives

$$x'(t) = r'(t)\cos\theta(t) - r(t)\theta'(t)\sin\theta(t), \quad y'(t) = r'(t)\sin\theta(t) + r(t)\theta'(t)\cos\theta(t)$$

[17] See O'NEILL for an introduction to elementary differential geometry.

[18] For example, the usual parametrizations of Bézier curves, which we discussed in Section 1.8, are polynomial functions in $\mathbb{R}^3$. This makes their computation relatively simple, which, in CAD, is desirable. But their arc length parametrizations are not only *not* polynomials, they are in fact usually impossible to calculate at all.

and so

$$
\begin{aligned}
x'(t)^2 + y'(t)^2 &= (r'(t)\cos\theta(t) - r(t)\theta'(t)\sin\theta(t))^2 + (r'(t)\sin\theta(t) + r(t)\theta'(t)\cos\theta(t))^2 \\
&= r'(t)^2(\cos^2\theta + \sin^2\theta) + r(t)^2\theta'(t)^2(\cos^2\theta + \sin^2\theta) \\
&\quad - 2r'(t)r(t)\theta'(t)\cos\theta\sin\theta + 2r'(t)r(t)\theta'(t)\cos\theta\sin\theta \\
&= r'(t)^2 + r(t)^2\theta'(t)^2 \quad \text{, and so} \\
L &= \int_a^b \sqrt{x'(t)^2 + y'(t)^2 + z'(t)^2}\,dt \\
&= \int_a^b \sqrt{r'(t)^2 + r(t)^2\theta'(t)^2 + z'(t)^2}\,dt
\end{aligned}
$$

QED

Example 1.44. Find the arc length L of the curve whose cylindrical coordinates are $r = e^t$, $\theta = t$ and $z = e^t$, for t over the interval $[0,1]$.

Solution: Since $r'(t) = e^t$, $\theta'(t) = 1$ and $z'(t) = e^t$, then

$$
\begin{aligned}
L &= \int_0^1 \sqrt{r'(t)^2 + r(t)^2\theta'(t)^2 + z'(t)^2}\,dt \\
&= \int_0^1 \sqrt{e^{2t} + e^{2t}(1) + e^{2t}}\,dt \\
&= \int_0^1 e^t\sqrt{3}\,dt = \sqrt{3}(e-1)
\end{aligned}
$$

Exercises

A

For Exercises 1-3, calculate the arc length of $\mathbf{f}(t)$ over the given interval.

1. $\mathbf{f}(t) = (3\cos 2t, 3\sin 2t, 3t)$ on $[0, \pi/2]$

2. $\mathbf{f}(t) = ((t^2+1)\cos t, (t^2+1)\sin t, 2\sqrt{2}t)$ on $[0,1]$

3. $\mathbf{f}(t) = (2\cos 3t, 2\sin 3t, 2t^{3/2})$ on $[0,1]$

4. Parametrize the curve from Exercise 1 by arc length.

5. Parametrize the curve from Exercise 3 by arc length.

B

6. Let $\mathbf{f}(t)$ be a differentiable curve such that $\mathbf{f}(t) \neq \mathbf{0}$ for all t. Show that

$$\frac{d}{dt}\left(\frac{\mathbf{f}(t)}{\|\mathbf{f}(t)\|}\right) = \frac{\mathbf{f}(t) \times (\mathbf{f}'(t) \times \mathbf{f}(t))}{\|\mathbf{f}(t)\|^3}.$$

Exercises 7-9 develop the moving frame field **T**, **N**, **B** at a point on a curve.

7. Let $\mathbf{f}(t)$ be a smooth curve such that $\mathbf{f}'(t) \neq \mathbf{0}$ for all t. Then we can define the *unit tangent vector* **T** by

$$\mathbf{T}(t) = \frac{\mathbf{f}'(t)}{\|\mathbf{f}'(t)\|}.$$

Show that

$$\mathbf{T}'(t) = \frac{\mathbf{f}'(t) \times (\mathbf{f}''(t) \times \mathbf{f}'(t))}{\|\mathbf{f}'(t)\|^3}.$$

8. Continuing Exercise 7, assume that $\mathbf{f}'(t)$ and $\mathbf{f}''(t)$ are not parallel. Then $\mathbf{T}'(t) \neq \mathbf{0}$ so we can define the *unit principal normal vector* **N** by

$$\mathbf{N}(t) = \frac{\mathbf{T}'(t)}{\|\mathbf{T}'(t)\|}.$$

Show that

$$\mathbf{N}(t) = \frac{\mathbf{f}'(t) \times (\mathbf{f}''(t) \times \mathbf{f}'(t))}{\|\mathbf{f}'(t)\| \, \|\mathbf{f}''(t) \times \mathbf{f}'(t)\|}.$$

9. Continuing Exercise 8, the *unit binormal vector* B is defined by

$$\mathbf{B}(t) = \mathbf{T}(t) \times \mathbf{N}(t).$$

Show that

$$\mathbf{B}(t) = \frac{\mathbf{f}'(t) \times \mathbf{f}''(t)}{\|\mathbf{f}'(t) \times \mathbf{f}''(t)\|}.$$

Note: The vectors $\mathbf{T}(t)$, $\mathbf{N}(t)$ and $\mathbf{B}(t)$ form a right-handed system of mutually perpendicular unit vectors (called *orthonormal vectors*) at each point on the curve $\mathbf{f}(t)$.

10. Continuing Exercise 9, the *curvature* κ is defined by

$$\kappa(t) = \frac{\|\mathbf{T}'(t)\|}{\|\mathbf{f}'(t)\|} = \frac{\|\mathbf{f}'(t) \times (\mathbf{f}''(t) \times \mathbf{f}'(t))\|}{\|\mathbf{f}'(t)\|^4}.$$

Show that

$$\kappa(t) = \frac{\|\mathbf{f}'(t) \times \mathbf{f}''(t)\|}{\|\mathbf{f}'(t)\|^3} \quad \text{and that} \quad \mathbf{T}'(t) = \|\mathbf{f}'(t)\| \, \kappa(t) \mathbf{N}(t).$$

Note: $\kappa(t)$ gives a sense of how "curved" the curve $\mathbf{f}(t)$ is at each point.

11. Find **T**, **N**, **B** and κ at each point of the helix $\mathbf{f}(t) = (\cos t, \sin t, t)$.

12. Show that the arc length L of a curve whose spherical coordinates are $\rho = \rho(t)$, $\theta = \theta(t)$ and $\phi = \phi(t)$ for t in an interval $[a,b]$ is

$$L = \int_a^b \sqrt{\rho'(t)^2 + (\rho(t)^2 \sin^2 \phi(t))\theta'(t)^2 + \rho(t)^2 \phi'(t)^2} \, dt.$$

2 Functions of Several Variables

2.1 Functions of Two or Three Variables

In Section 1.8 we discussed vector-valued functions of a single real variable. We will now examine real-valued functions of a point (or vector) in $\mathbb{R}^2$ or $\mathbb{R}^3$. For the most part these functions will be defined on sets of points in $\mathbb{R}^2$, but there will be times when we will use points in $\mathbb{R}^3$, and there will also be times when it will be convenient to think of the points as vectors (or terminal points of vectors).

A **real-valued function** f defined on a subset D of $\mathbb{R}^2$ is a rule that assigns to each point (x, y) in D a real number $f(x, y)$. The largest possible set D in $\mathbb{R}^2$ on which f is defined is called the **domain** of f, and the **range** of f is the set of all real numbers $f(x, y)$ as (x, y) varies over the domain D. A similar definition holds for functions $f(x, y, z)$ defined on points (x, y, z) in $\mathbb{R}^3$.

Example 2.1. The domain of the function

$$f(x, y) = xy$$

is all of $\mathbb{R}^2$, and the range of f is all of $\mathbb{R}$.

Example 2.2. The domain of the function

$$f(x, y) = \frac{1}{x - y}$$

is all of $\mathbb{R}^2$ except the points (x, y) for which $x = y$. That is, the domain is the set $D = \{(x, y) : x \neq y\}$. The range of f is all real numbers except 0.

Example 2.3. The domain of the function

$$f(x, y) = \sqrt{1 - x^2 - y^2}$$

is the set $D = \{(x, y) : x^2 + y^2 \leq 1\}$, since the quantity inside the square root is nonnegative if and only if $1 - (x^2 + y^2) \geq 0$. We see that D consists of all points on and inside the unit circle in $\mathbb{R}^2$ (D is sometimes called the *closed unit disk*). The range of f is the interval $[0, 1]$ in $\mathbb{R}$.

Example 2.4. The domain of the function

$$f(x,y,z) = e^{x+y-z}$$

is all of $\mathbb{R}^3$, and the range of f is all positive real numbers.

A function $f(x,y)$ defined in $\mathbb{R}^2$ is often written as $z = f(x,y)$, as was mentioned in Section 1.1, so that the **graph** of $f(x,y)$ is the set $\{(x,y,z) : z = f(x,y)\}$ in $\mathbb{R}^3$. So we see that this graph is a surface in $\mathbb{R}^3$, since it satisfies an equation of the form $F(x,y,z) = 0$ (namely, $F(x,y,z) = f(x,y) - z$). The traces of this surface in the planes $z = c$, where c varies over $\mathbb{R}$, are called the **level curves** of the function. Equivalently, the level curves are the solution sets of the equations $f(x,y) = c$, for c in $\mathbb{R}$. Level curves are often projected onto the xy-plane to give an idea of the various "elevation" levels of the surface (as is done in topography).

Example 2.5. The graph of the function

$$f(x,y) = \frac{\sin\sqrt{x^2+y^2}}{\sqrt{x^2+y^2}}$$

is shown below. Note that the level curves (shown both on the surface and projected onto the xy-plane) are groups of concentric circles.

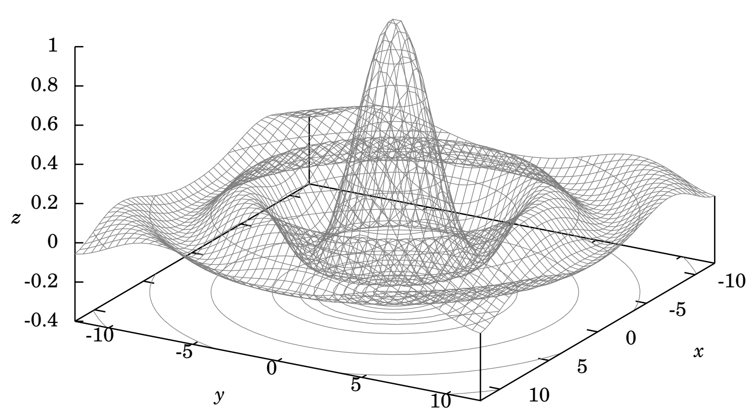


Figure 2.1.1 The function $f(x,y) = \frac{\sin\sqrt{x^2+y^2}}{\sqrt{x^2+y^2}}$

You may be wondering what happens to the function in Example 2.5 at the point $(x,y) = (0,0)$, since both the numerator and denominator are 0 at that point. The function is not defined at $(0,0)$, but the *limit* of the function exists (and equals 1) as (x,y) *approaches* $(0,0)$. We will now state explicitly what is meant by the limit of a function of two variables.

Definition 2.1. Let (a,b) be a point in $\mathbb{R}^2$, and let $f(x,y)$ be a real-valued function defined on some set containing (a,b) (but not necessarily defined at (a,b) itself). Then we say that the **limit** of $f(x,y)$ equals L as (x,y) approaches (a,b), written as

$$\lim_{(x,y)\to(a,b)} f(x,y) = L \,, \tag{2.1}$$

if given any $\epsilon > 0$, there exists a $\delta > 0$ such that

$$|f(x,y) - L| < \epsilon \quad \text{whenever} \quad 0 < \sqrt{(x-a)^2 + (y-b)^2} < \delta.$$

A similar definition can be made for functions of three variables. The idea behind the above definition is that the values of $f(x,y)$ can get arbitrarily close to L (i.e. within ϵ of L) if we pick (x,y) sufficiently close to (a,b) (i.e. inside a circle centered at (a,b) with some sufficiently small radius δ).

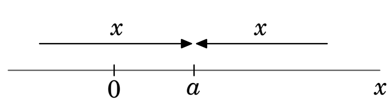
If you recall the "epsilon-delta" proofs of limits of real-valued functions of a single variable, you may remember how awkward they can be, and how they can usually only be done easily for simple functions. In general, the multivariable cases are at least equally awkward to go through, so we will not bother with such proofs. Instead, we will simply state that when the function $f(x,y)$ is given by a single formula and is defined at the point (a,b) (e.g. is not some indeterminate form like 0/0) then you can just substitute $(x,y) = (a,b)$ into the formula for $f(x,y)$ to find the limit.

Example 2.6.

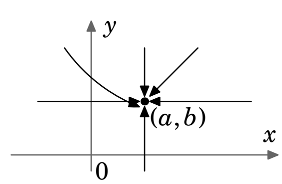
$$\lim_{(x,y)\to(1,2)} \frac{xy}{x^2+y^2} = \frac{(1)(2)}{1^2+2^2} = \frac{2}{5}$$

since $f(x,y) = \frac{xy}{x^2+y^2}$ is properly defined at the point $(1,2)$.

The major difference between limits in one variable and limits in two or more variables has to do with how a point is approached. In the single-variable case, the statement "$x \to a$" means that x gets closer to the value a from two possible directions along the real number line (see Figure 2.1.2(a)). In two dimensions, however, (x,y) can approach a point (a,b) along an infinite number of paths (see Figure 2.1.2(b)).



(a) $x \to a$ in $\mathbb{R}$



(b) $(x,y) \to (a,b)$ in $\mathbb{R}^2$

Figure 2.1.2 "Approaching" a point in different dimensions

Example 2.7.

$$\lim_{(x,y)\to(0,0)} \frac{xy}{x^2+y^2} \quad \text{does not exist}$$

Note that we can not simply substitute $(x, y) = (0,0)$ into the function, since doing so gives an indeterminate form 0/0. To show that the limit does not exist, we will show that the function approaches different values as (x, y) approaches $(0,0)$ along different paths in $\mathbb{R}^2$. To see this, suppose that $(x, y) \to (0,0)$ along the positive x-axis, so that $y = 0$ along that path. Then

$$f(x,y) = \frac{xy}{x^2+y^2} = \frac{x0}{x^2+0^2} = 0$$

along that path (since $x > 0$ in the denominator). But if $(x, y) \to (0,0)$ along the straight line $y = x$ through the origin, for $x > 0$, then we see that

$$f(x,y) = \frac{xy}{x^2+y^2} = \frac{x^2}{x^2+x^2} = \frac{1}{2}$$

which means that $f(x, y)$ approaches different values as $(x, y) \to (0,0)$ along different paths. Hence the limit does not exist.

Limits of real-valued multivariable functions obey the same algebraic rules as in the single-variable case, as shown in the following theorem, which we state without proof.

Theorem 2.1. Suppose that $\lim\limits_{(x,y)\to(a,b)} f(x,y)$ and $\lim\limits_{(x,y)\to(a,b)} g(x,y)$ both exist, and that k is some scalar. Then:

(a) $\displaystyle \lim_{(x,y)\to(a,b)} [f(x,y) \pm g(x,y)] = \left[\lim_{(x,y)\to(a,b)} f(x,y) \right] \pm \left[\lim_{(x,y)\to(a,b)} g(x,y) \right]$

(b) $\displaystyle \lim_{(x,y)\to(a,b)} k\,f(x,y) = k \left[\lim_{(x,y)\to(a,b)} f(x,y) \right]$

(c) $\displaystyle \lim_{(x,y)\to(a,b)} [f(x,y)g(x,y)] = \left[\lim_{(x,y)\to(a,b)} f(x,y) \right] \left[\lim_{(x,y)\to(a,b)} g(x,y) \right]$

(d) $\displaystyle \lim_{(x,y)\to(a,b)} \frac{f(x,y)}{g(x,y)} = \frac{\lim\limits_{(x,y)\to(a,b)} f(x,y)}{\lim\limits_{(x,y)\to(a,b)} g(x,y)} \quad \text{if} \lim_{(x,y)\to(a,b)} g(x,y) \neq 0$

(e) If $|f(x,y) - L| \le g(x,y)$ for all (x, y) and if $\lim\limits_{(x,y)\to(a,b)} g(x,y) = 0$, then $\lim\limits_{(x,y)\to(a,b)} f(x,y) = L$.

Note that in part (e), it suffices to have $|f(x,y) - L| \le g(x,y)$ for all (x, y) “sufficiently close” to (a, b) (but excluding (a, b) itself).

Example 2.8. Show that

$$\lim_{(x,y)\to(0,0)} \frac{y^4}{x^2+y^2} = 0.$$

Since substituting $(x,y) = (0,0)$ into the function gives the indeterminate form 0/0, we need an alternate method for evaluating this limit. We will use Theorem 2.1(e). First, notice that $y^4 = \left(\sqrt{y^2}\right)^4$ and so $0 \le y^4 \le \left(\sqrt{x^2+y^2}\right)^4$ for all (x,y). But $\left(\sqrt{x^2+y^2}\right)^4 = (x^2+y^2)^2$. Thus, for all $(x,y) \ne (0,0)$ we have

$$\left|\frac{y^4}{x^2+y^2}\right| \le \frac{(x^2+y^2)^2}{x^2+y^2} = x^2+y^2 \to 0 \quad \text{as} \quad (x,y)\to(0,0).$$

Therefore $\displaystyle\lim_{(x,y)\to(0,0)} \frac{y^4}{x^2+y^2} = 0$.

Continuity can be defined similarly as in the single-variable case.

Definition 2.2. A real-valued function $f(x,y)$ with domain D in $\mathbb{R}^2$ is **continuous** at the point (a,b) in D if $\displaystyle\lim_{(x,y)\to(a,b)} f(x,y) = f(a,b)$. We say that $f(x,y)$ is a **continuous function** if it is continuous at every point in its domain D.

Unless indicated otherwise, you can assume that all the functions we deal with are continuous. In fact, we can modify the function from Example 2.8 so that it is continuous on all of $\mathbb{R}^2$.

Example 2.9. Define a function $f(x,y)$ on all of $\mathbb{R}^2$ as follows:

$$f(x,y) = \begin{cases} 0 & \text{if } (x,y) = (0,0) \\ \dfrac{y^4}{x^2+y^2} & \text{if } (x,y) \ne (0,0) \end{cases}$$

Then $f(x,y)$ is well-defined for all (x,y) in $\mathbb{R}^2$ (i.e. there are no indeterminate forms for any (x,y)), and we see that

$$\lim_{(x,y)\to(a,b)} f(x,y) = \frac{b^4}{a^2+b^2} = f(a,b) \quad \text{for } (a,b) \ne (0,0).$$

So since

$$\lim_{(x,y)\to(0,0)} f(x,y) = 0 = f(0,0) \quad \text{by Example 2.8,}$$

then $f(x,y)$ is continuous on all of $\mathbb{R}^2$.

Exercises

A

For Exercises 1-6, state the domain and range of the given function.

1. $f(x,y) = x^2 + y^2 - 1$

2. $f(x,y) = \dfrac{1}{x^2 + y^2}$

3. $f(x,y) = \sqrt{x^2 + y^2 - 4}$

4. $f(x,y) = \dfrac{x^2 + 1}{y}$

5. $f(x,y,z) = \sin(xyz)$

6. $f(x,y,z) = \sqrt{(x-1)(yz-1)}$

For Exercises 7-18, evaluate the given limit.

7. $\displaystyle\lim_{(x,y)\to(0,0)} \cos(xy)$

8. $\displaystyle\lim_{(x,y)\to(0,0)} e^{xy}$

9. $\displaystyle\lim_{(x,y)\to(0,0)} \frac{x^2 - y^2}{x^2 + y^2}$

10. $\displaystyle\lim_{(x,y)\to(0,0)} \frac{xy^2}{x^2 + y^4}$

11. $\displaystyle\lim_{(x,y)\to(1,-1)} \frac{x^2 - 2xy + y^2}{x - y}$

12. $\displaystyle\lim_{(x,y)\to(0,0)} \frac{xy^2}{x^2 + y^2}$

13. $\displaystyle\lim_{(x,y)\to(1,1)} \frac{x^2 - y^2}{x - y}$

14. $\displaystyle\lim_{(x,y)\to(0,0)} \frac{x^2 - 2xy + y^2}{x - y}$

15. $\displaystyle\lim_{(x,y)\to(0,0)} \frac{y^4 \sin(xy)}{x^2 + y^2}$

16. $\displaystyle\lim_{(x,y)\to(0,0)} (x^2 + y^2)\cos\left(\frac{1}{xy}\right)$

17. $\displaystyle\lim_{(x,y)\to(0,0)} \frac{x}{y}$

18. $\displaystyle\lim_{(x,y)\to(0,0)} \cos\left(\frac{1}{xy}\right)$

B

19. Show that $f(x,y) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}$, for $\sigma > 0$, is constant on the circle of radius $r > 0$ centered at the origin. This function is called a *Gaussian blur*, and is used as a filter in image processing software to produce a “blurred” effect.

20. Suppose that $f(x,y) \le f(y,x)$ for all (x,y) in $\mathbb{R}^2$. Show that $f(x,y) = f(y,x)$ for all (x,y) in $\mathbb{R}^2$.

21. Use the substitution $r = \sqrt{x^2 + y^2}$ to show that

$$\lim_{(x,y)\to(0,0)} \frac{\sin\sqrt{x^2 + y^2}}{\sqrt{x^2 + y^2}} = 1 .$$

(*Hint: You will need to use L'Hôpital's Rule for single-variable limits.*)

C

22. Prove Theorem 2.1(a) in the case of addition. (*Hint: Use Definition 2.1.*)

23. Prove Theorem 2.1(b).

2.2 Partial Derivatives

Now that we have an idea of what functions of several variables are, and what a limit of such a function is, we can start to develop an idea of a derivative of a function of two or more variables. We will start with the notion of a *partial derivative*.

Definition 2.3. Let $f(x,y)$ be a real-valued function with domain D in $\mathbb{R}^2$, and let (a,b) be a point in D. Then the **partial derivative of f at (a,b) with respect to x**, denoted by $\frac{\partial f}{\partial x}(a,b)$, is defined as

$$\frac{\partial f}{\partial x}(a,b) = \lim_{h\to 0} \frac{f(a+h,b)-f(a,b)}{h} \tag{2.2}$$

and the **partial derivative of f at (a,b) with respect to y**, denoted by $\frac{\partial f}{\partial y}(a,b)$, is defined as

$$\frac{\partial f}{\partial y}(a,b) = \lim_{h\to 0} \frac{f(a,b+h)-f(a,b)}{h}\ . \tag{2.3}$$

Note: The symbol ∂ is pronounced "del".[1]

Recall that the derivative of a function $f(x)$ can be interpreted as the rate of change of that function in the (positive) x direction. From the definitions above, we can see that the partial derivative of a function $f(x,y)$ with respect to x or y is the rate of change of $f(x,y)$ in the (positive) x or y direction, respectively. What this means is that the partial derivative of a function $f(x,y)$ with respect to x can be calculated by treating the y *variable* as a *constant*, and then simply differentiating $f(x,y)$ as if it were a function of x alone, using the usual rules from single-variable calculus. Likewise, the partial derivative of $f(x,y)$ with respect to y is obtained by treating the x variable as a constant and then differentiating $f(x,y)$ as if it were a function of y alone.

Example 2.10. Find $\frac{\partial f}{\partial x}(x,y)$ and $\frac{\partial f}{\partial y}(x,y)$ for the function $f(x,y) = x^2y + y^3$.

Solution: Treating y as a constant and differentiating $f(x,y)$ with respect to x gives

$$\frac{\partial f}{\partial x}(x,y) = 2xy$$

and treating x as a constant and differentiating $f(x,y)$ with respect to y gives

$$\frac{\partial f}{\partial y}(x,y) = x^2 + 3y^2\ .$$

---

[1] It is not a Greek letter. The symbol was first used by the mathematicians A. Clairaut and L. Euler around 1740, to distinguish it from the letter d used for the "usual" derivative.

We will often simply write $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ instead of $\frac{\partial f}{\partial x}(x,y)$ and $\frac{\partial f}{\partial y}(x,y)$.

Example 2.11. Find $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ for the function $f(x,y) = \frac{\sin(xy^2)}{x^2+1}$.

Solution: Treating y as a constant and differentiating $f(x,y)$ with respect to x gives

$$\frac{\partial f}{\partial x} = \frac{(x^2+1)(y^2\cos(xy^2)) - (2x)\sin(xy^2)}{(x^2+1)^2}$$

and treating x as a constant and differentiating $f(x,y)$ with respect to y gives

$$\frac{\partial f}{\partial y} = \frac{2xy\cos(xy^2)}{x^2+1}\;.$$

Since both $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ are themselves functions of x and y, we can take *their* partial derivatives with respect to x and y. This yields the *higher-order partial derivatives*:

$$\begin{aligned}
\frac{\partial^2 f}{\partial x^2} &= \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right) & \frac{\partial^2 f}{\partial y^2} &= \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial y}\right)\\
\frac{\partial^2 f}{\partial y\,\partial x} &= \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right) & \frac{\partial^2 f}{\partial x\,\partial y} &= \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right)\\
\frac{\partial^3 f}{\partial x^3} &= \frac{\partial}{\partial x}\left(\frac{\partial^2 f}{\partial x^2}\right) & \frac{\partial^3 f}{\partial y^3} &= \frac{\partial}{\partial y}\left(\frac{\partial^2 f}{\partial y^2}\right)\\
\frac{\partial^3 f}{\partial y\,\partial x^2} &= \frac{\partial}{\partial y}\left(\frac{\partial^2 f}{\partial x^2}\right) & \frac{\partial^3 f}{\partial x\,\partial y^2} &= \frac{\partial}{\partial x}\left(\frac{\partial^2 f}{\partial y^2}\right)\\
\frac{\partial^3 f}{\partial y^2\,\partial x} &= \frac{\partial}{\partial y}\left(\frac{\partial^2 f}{\partial y\,\partial x}\right) & \frac{\partial^3 f}{\partial x^2\,\partial y} &= \frac{\partial}{\partial x}\left(\frac{\partial^2 f}{\partial x\,\partial y}\right)\\
\frac{\partial^3 f}{\partial x\,\partial y\,\partial x} &= \frac{\partial}{\partial x}\left(\frac{\partial^2 f}{\partial y\,\partial x}\right) & \frac{\partial^3 f}{\partial y\,\partial x\,\partial y} &= \frac{\partial}{\partial y}\left(\frac{\partial^2 f}{\partial x\,\partial y}\right)\\
&\vdots
\end{aligned}$$

Example 2.12. Find the partial derivatives $\frac{\partial f}{\partial x}$, $\frac{\partial f}{\partial y}$, $\frac{\partial^2 f}{\partial x^2}$, $\frac{\partial^2 f}{\partial y^2}$, $\frac{\partial^2 f}{\partial y\,\partial x}$ and $\frac{\partial^2 f}{\partial x\,\partial y}$ for the function $f(x,y) = e^{x^2 y} + xy^3$.

Solution: Proceeding as before, we have

$$\frac{\partial f}{\partial x} = 2xye^{x^2y} + y^3 \qquad \frac{\partial f}{\partial y} = x^2e^{x^2y} + 3xy^2$$

$$\frac{\partial^2 f}{\partial x^2} = \frac{\partial}{\partial x}(2xye^{x^2y} + y^3) \qquad \frac{\partial^2 f}{\partial y^2} = \frac{\partial}{\partial y}(x^2e^{x^2y} + 3xy^2)$$

$$= 2ye^{x^2y} + 4x^2y^2e^{x^2y} \qquad = x^4e^{x^2y} + 6xy$$

$$\frac{\partial^2 f}{\partial y\,\partial x} = \frac{\partial}{\partial y}(2xye^{x^2y} + y^3) \qquad \frac{\partial^2 f}{\partial x\,\partial y} = \frac{\partial}{\partial x}(x^2e^{x^2y} + 3xy^2)$$

$$= 2xe^{x^2y} + 2x^3ye^{x^2y} + 3y^2 \qquad = 2xe^{x^2y} + 2x^3ye^{x^2y} + 3y^2$$

Higher-order partial derivatives that are taken with respect to different variables, such as $\frac{\partial^2 f}{\partial y\,\partial x}$ and $\frac{\partial^2 f}{\partial x\,\partial y}$, are called **mixed partial derivatives**. Notice in the above example that $\frac{\partial^2 f}{\partial y\,\partial x} = \frac{\partial^2 f}{\partial x\,\partial y}$. It turns that this will usually be the case. Specifically, whenever both $\frac{\partial^2 f}{\partial y\,\partial x}$ and $\frac{\partial^2 f}{\partial x\,\partial y}$ are continuous at a point (a,b), then they are equal at that point.[2] All the functions we will deal with will have continuous partial derivatives of all orders, so you can assume in the remainder of the text that

$$\frac{\partial^2 f}{\partial y\,\partial x} = \frac{\partial^2 f}{\partial x\,\partial y} \quad \text{for all } (x,y) \text{ in the domain of } f.$$

In other words, it doesn't matter in which order you take partial derivatives. This applies even to mixed partial derivatives of order 3 or higher.

The notation for partial derivatives varies. All of the following are equivalent:

$$\frac{\partial f}{\partial x} : \ f_x(x,y)\,, \quad f_1(x,y)\,, \quad D_x(x,y)\,, \quad D_1(x,y)$$

$$\frac{\partial f}{\partial y} : \ f_y(x,y)\,, \quad f_2(x,y)\,, \quad D_y(x,y)\,, \quad D_2(x,y)$$

$$\frac{\partial^2 f}{\partial x^2} : \ f_{xx}(x,y)\,, \quad f_{11}(x,y)\,, \quad D_{xx}(x,y)\,, \quad D_{11}(x,y)$$

$$\frac{\partial^2 f}{\partial y^2} : \ f_{yy}(x,y)\,, \quad f_{22}(x,y)\,, \quad D_{yy}(x,y)\,, \quad D_{22}(x,y)$$

$$\frac{\partial^2 f}{\partial y\,\partial x} : \ f_{xy}(x,y)\,, \quad f_{12}(x,y)\,, \quad D_{xy}(x,y)\,, \quad D_{12}(x,y)$$

$$\frac{\partial^2 f}{\partial x\,\partial y} : \ f_{yx}(x,y)\,, \quad f_{21}(x,y)\,, \quad D_{yx}(x,y)\,, \quad D_{21}(x,y)$$

[2] See pp. 214-216 in TAYLOR and MANN for a proof.

Exercises

A

For Exercises 1-16, find $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$.

1. $f(x,y) = x^2 + y^2$

2. $f(x,y) = \cos(x+y)$

3. $f(x,y) = \sqrt{x^2 + y + 4}$

4. $f(x,y) = \dfrac{x+1}{y+1}$

5. $f(x,y) = e^{xy} + xy$

6. $f(x,y) = x^2 - y^2 + 6xy + 4x - 8y + 2$

7. $f(x,y) = x^4$

8. $f(x,y) = x + 2y$

9. $f(x,y) = \sqrt{x^2 + y^2}$

10. $f(x,y) = \sin(x+y)$

11. $f(x,y) = \sqrt[3]{x^2 + y + 4}$

12. $f(x,y) = \dfrac{xy+1}{x+y}$

13. $f(x,y) = e^{-(x^2+y^2)}$

14. $f(x,y) = \ln(xy)$

15. $f(x,y) = \sin(xy)$

16. $f(x,y) = \tan(x+y)$

For Exercises 17-26, find $\frac{\partial^2 f}{\partial x^2}$, $\frac{\partial^2 f}{\partial y^2}$ and $\frac{\partial^2 f}{\partial y \partial x}$ (use Exercises 1-8, 14, 15).

17. $f(x,y) = x^2 + y^2$

18. $f(x,y) = \cos(x+y)$

19. $f(x,y) = \sqrt{x^2 + y + 4}$

20. $f(x,y) = \dfrac{x+1}{y+1}$

21. $f(x,y) = e^{xy} + xy$

22. $f(x,y) = x^2 - y^2 + 6xy + 4x - 8y + 2$

23. $f(x,y) = x^4$

24. $f(x,y) = x + 2y$

25. $f(x,y) = \ln(xy)$

26. $f(x,y) = \sin(xy)$

B

27. Show that the function $f(x,y) = \sin(x+y) + \cos(x-y)$ satisfies the *wave equation*

$$\frac{\partial^2 f}{\partial x^2} - \frac{\partial^2 f}{\partial y^2} = 0 \;.$$

The wave equation is an example of a *partial differential equation*.

28. Let u and v be twice-differentiable functions of a single variable, and let $c \neq 0$ be a constant. Show that $f(x,y) = u(x+cy) + v(x-cy)$ is a solution of the *general one-dimensional wave equation*[3]

$$\frac{\partial^2 f}{\partial x^2} - \frac{1}{c^2}\frac{\partial^2 f}{\partial y^2} = 0 \;.$$

[3]Conversely, it turns out that *any* solution must be of this form. See Ch. 1 in WEINBERGER.